# **Report Outlining GHG Emissions in the UK**

# Introduction and Data Overview

This report investigates greenhouse gas (GHG) emissions in the UK over time and forecasts future emissions to assess if the UK is on track to achieve its net zero target by 2050. Emissions data from the ONS atmospheric emissions dataset is used to build a time series model predicting emissions over the next 5 years.

The dataset (current edition) contains annual GHG emissions amounts broken down by industry category from 1990-2021, with a total emission across all industries. It provides an industry grouping as well as economic sectors. I have focused on the ‘GHG total’ sheet, since other sheets do not add any value to the goal of this exercise, although may provide further insight into which gases contribute the most to gas emission. Moreover, the breakdown of industries into sectors will also be ignored, with the similar reasoning as stated for gases, as it also is combined to showcase the total GHG emission, which is already simplified and presented in the first table.

# Data Cleaning

The data was pre-processed to handle any missing values or erroneous data points, the following steps were taken to achieve clean data:

* Dropped NaN rows and unnamed columns.
* Converted data types (during modelling processes) to prepare data for modelling.
* Renamed column to “industries” for more clarification.
* Manually deleted cells in Excel not involved with data. This step will normally be completed in Python, however since the cells in Excel were not correlated with data, and would not change the data itself, I have deleted cells for simplicity.
* Deleted the breakdown into sectors’ data to avoid complexity.

# Exploratory Data Analysis

During this phase, I have decided to investigate the data further to gain greater understanding of the data at hand, while also finding any raises for concern, trends, and any numerical values that may contribute to the overall analysis of the data. Here are the patterns and results I have discovered:

* Total GHG has been constantly declining since 1990 (Figure 1), with a total mean decrease of -40.35% from 1990-2021, and as a result, variability decreased from a standard deviation of 182142.60 in 1990 to 108276.17 in 2021.
* Dataset reveals notable percentage changes in emission between specific years. The most substantial decrease occurred during 2019-2020 (Figure 2), with an overall drop of -11.26%, this period resonates with the introduction of the new legislation in 2019, displaying a substantial impact of reduction of total GHG emission.
* Certain industries including construction, wholesale and retail trade, repair of motor vehicles, accommodation, and food service activities, exhibited slight increases in overall emission from 1990-2021, this growth may be attributed to growth in the industry, changed in production processes, or increased demand.
* Consumer expenditure had the third highest average emissions output, but only an approximate 6% decrease between 1990 and 2021 – lowest change among all industries. It also had the highest emission level in 2021, this stable and small reduction could be attributed to increasing usage by households over the years, counteracting emission goals.

A graph showing the growth of greenhouse gas prices

Description automatically generated

Figure Showing Gradual Decline of GHG Emission Over Time

A graph showing the growth of a greenhouse effect

Description automatically generated

Figure Showing Percentage Change in Emission Between Each Year

# Model Building

Using PyCaret, I was able to identify the best model for this specific dataset, since the data does not have any seasonality, and also contains white noise, models such as ARIMA, XGBoost, and Facebook Prophet will not work, or will provide insignificant results with low accuracy scores, considering the data is fairly small, and only one variable is used to predict emission (total greenhouse emission), a complex model would not be needed, thus two models were used, ETS and a regression model. These models were used to compare one another to find the model with best accuracy score, due to time constraints, with the project not requiring taking over a day, comparing more models would be unnecessary, especially when such a high accuracy score was achieved, the best model was shown to be ETS. The ETS model was fit on the emission data from 1990-2016, predictions were generated for 2017-2021, as well as plotted a line graph to display prediction of each from 1990 (Figure 3) and compared to the actual emission during those years. The model achieved a low MAPE of 2.47%, therefore an accuracy score of ~97.5%, an RMSE score of 16714.8465, since the accuracy was so high, a finalised model was made to forecast the next 5 years, presenting a gradual decline in UK GHG emissions (Figure 4). This was then further taken to forecast every year until 2050, where it is predicted that by 2050, there would be an output of -204.67, due to RMSE score, this may be further delayed to 2051, however, showing net zero goal is on track to being achieved.

The linear regression model was also tested to see how it stands against the ETS model, while with impressive performance, it only achieved an MAPE score of 4.0% and a higher MAE, not continuing into the forecast prediction since they will not be as accurate as the ETS model.

A graph with blue and orange lines

Description automatically generated

Figure Showing Prediction of Model Against Actual Values

A graph showing the growth of a stock market

Description automatically generated

Figure Showing Forecast of Next 5 Years Predicted by Model

# Summary & Implications

The analysis of the UK GHG emission unveils a consistent decline since 1990, reflecting an overall decrease of -40.35%. Events such as introduction of legislation in 2019, boosted the acceleration of a reduction of GHG emission. The findings suggest a positive trajectory towards the UK’s net zero target. Industries experiencing growth in emission may benefit from direct intervention, particularly consumer expenditure, creating necessary strategies to encourage sustainable consumption.

# Model Limitations & Improvement

The ETS model, while simple and accurate, has limitations. It assumes a linear trend, potentially oversimplifying complex dynamics. Future uses of the ETS model could incorporate additional variables like GDP for enhanced accuracy and refinement of the model, capturing real world dynamics. It could also test different seasonality timestamps and use ensemble methods, combining with other models to gain further insights.